



Incorporating Contextual and Syntactic Structures Improves Semantic Similarity Modeling



Linqing Liu, Wei Yang, Jinfeng Rao, Raphael Tang, and Jimmy Lin



Semantic Textual Similarity

Similarity score: 

- - Some people are walking
 - People are walking 
 - A group of scouts are camping in the grass 

- - A group of people is on a beach
 - A group of people is near the sea 
 - A group of people is on a mountain 

Current Models for Sentence Pair Modeling

- Shortcut-stacked Sentence Encoders (**SSE**) [Nie and Bansal, 2017]
- Decomposable Attention Model (**DecAtt**) [Parikh et al., 2016]
- Bi-LSTM Max-pooling Network (**InferSent**) [Conneau et al., 2017]
- Enhanced Sequential Inference Model (**ESIM**) [Chen et al., 2017]
- Pairwise Word Interaction model (**PWIM**) [He and Lin, 2016]

Current Models for Sentence Pair Modeling

Components

**Sentence Pair Interaction &
Attention Mechanism**

Sentence Representation

Current Models for Sentence Pair Modeling

Components

**Sentence Pair Interaction &
Attention Mechanism**

**Encoding Contextual
Information by LSTM**

**Incorporation of Syntactic
Parsing Information**

Current Models for Sentence Pair Modeling

	SSE	DecAtt	Infersent	ESIM	PWIM
Components					
Sentence Pair Interaction & Attention Mechanism		✓		✓	✓
Encoding Contextual Information by LSTM					
Incorporation of Syntactic Parsing Information					

Current Models for Sentence Pair Modeling

	SSE	DecAtt	Infersent	ESIM	PWIM
Components					
Sentence Pair Interaction & Attention Mechanism		✓		✓	✓
Encoding Contextual Information by LSTM	✓		✓	✓	✓
Incorporation of Syntactic Parsing Information					

Current Models for Sentence Pair Modeling

	SSE	DecAtt	Infersent	ESIM	PWIM
Components					
Sentence Pair Interaction & Attention Mechanism		✓		✓	✓
Encoding Contextual Information by LSTM	✓		✓	✓	✓
Incorporation of Syntactic Parsing Information				✓	

Current Models for Sentence Pair Modeling

	SSE	DecAtt	Infersent	ESIM	PWIM
Components					
Sentence Pair Interaction & Attention Mechanism	✓	✓	✓	✓	✓
Encoding Contextual Information by LSTM	✓	✓	✓	✓	✓
Incorporation of Syntactic Parsing Information				✓	

Does **Syntactic Structure** help Sentence Pair Modeling? ✓

Current Models for Sentence Pair Modeling

SSE

DecAtt

Infersent

ESIM

PWIM

Components

Sentence Pair Interaction &
Attention Mechanism

Encoding Contextual
Information by LSTM

Incorporation of Syntactic
Parsing Information

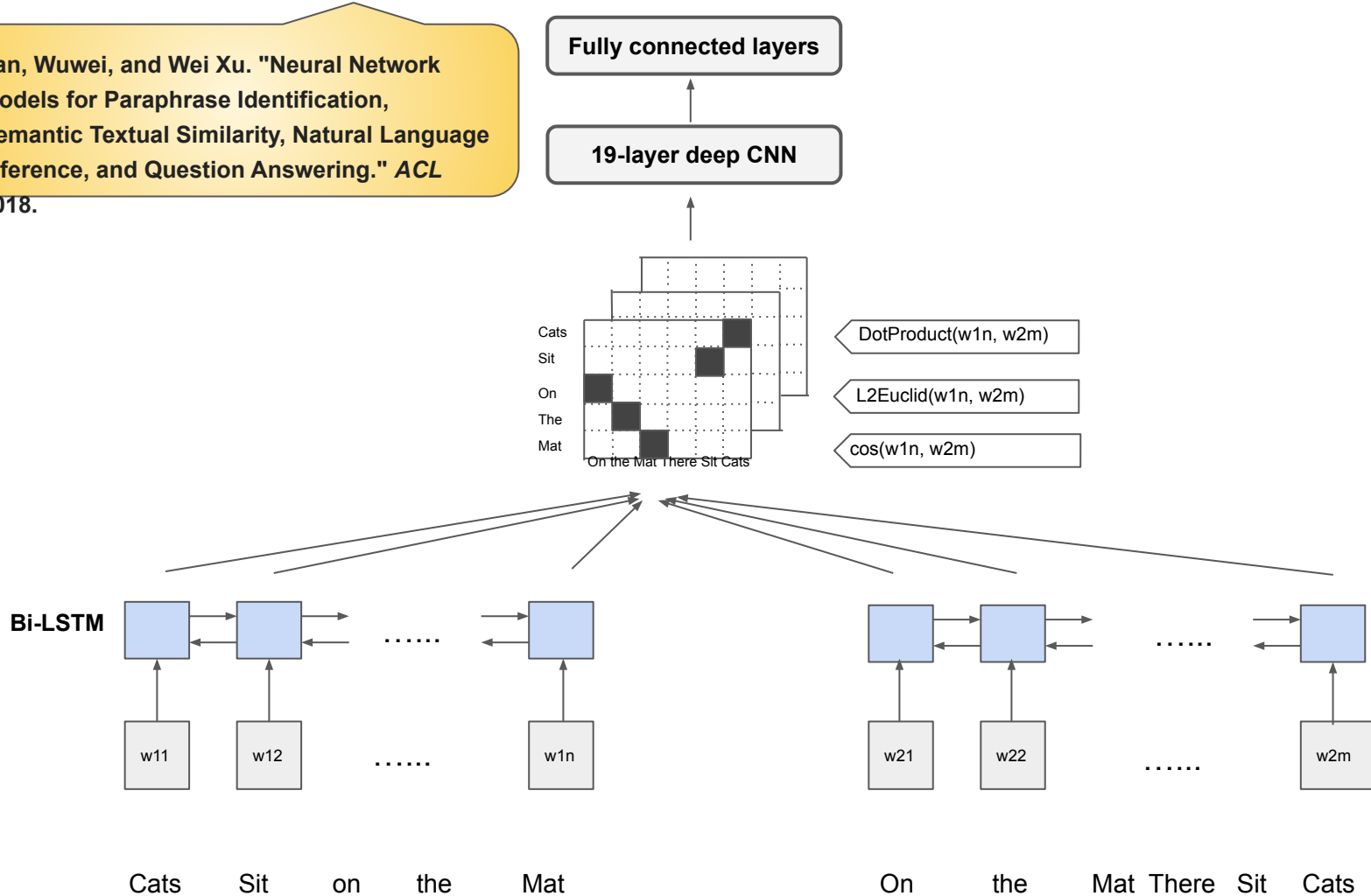
Do

**Contextual and Syntactic
Structures**

help sentence Pair Modeling?

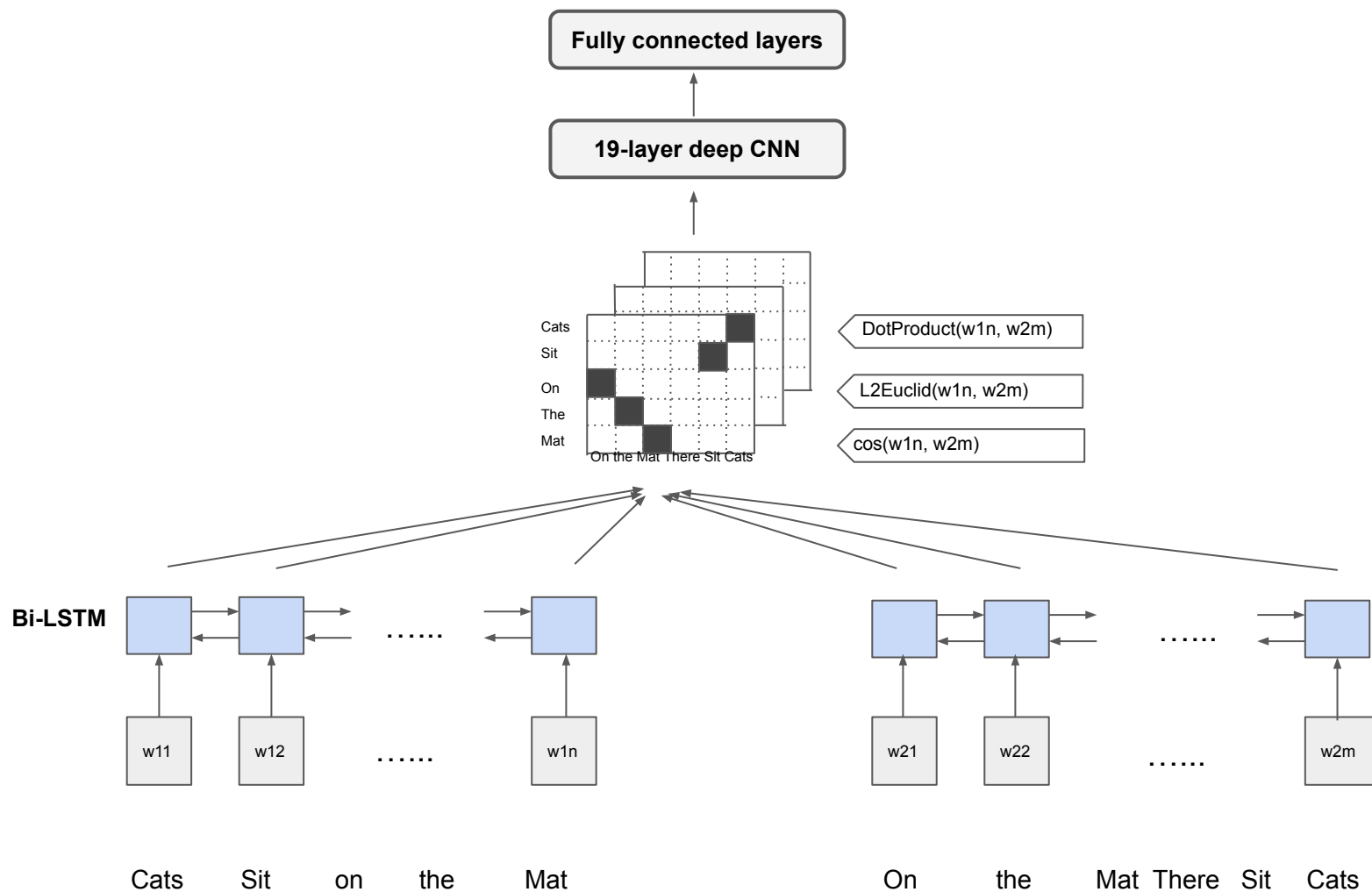
Pairwise Word Interaction Model (PWIM)

Lan, Wuwei, and Wei Xu. "Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering." *ACL* 2018.

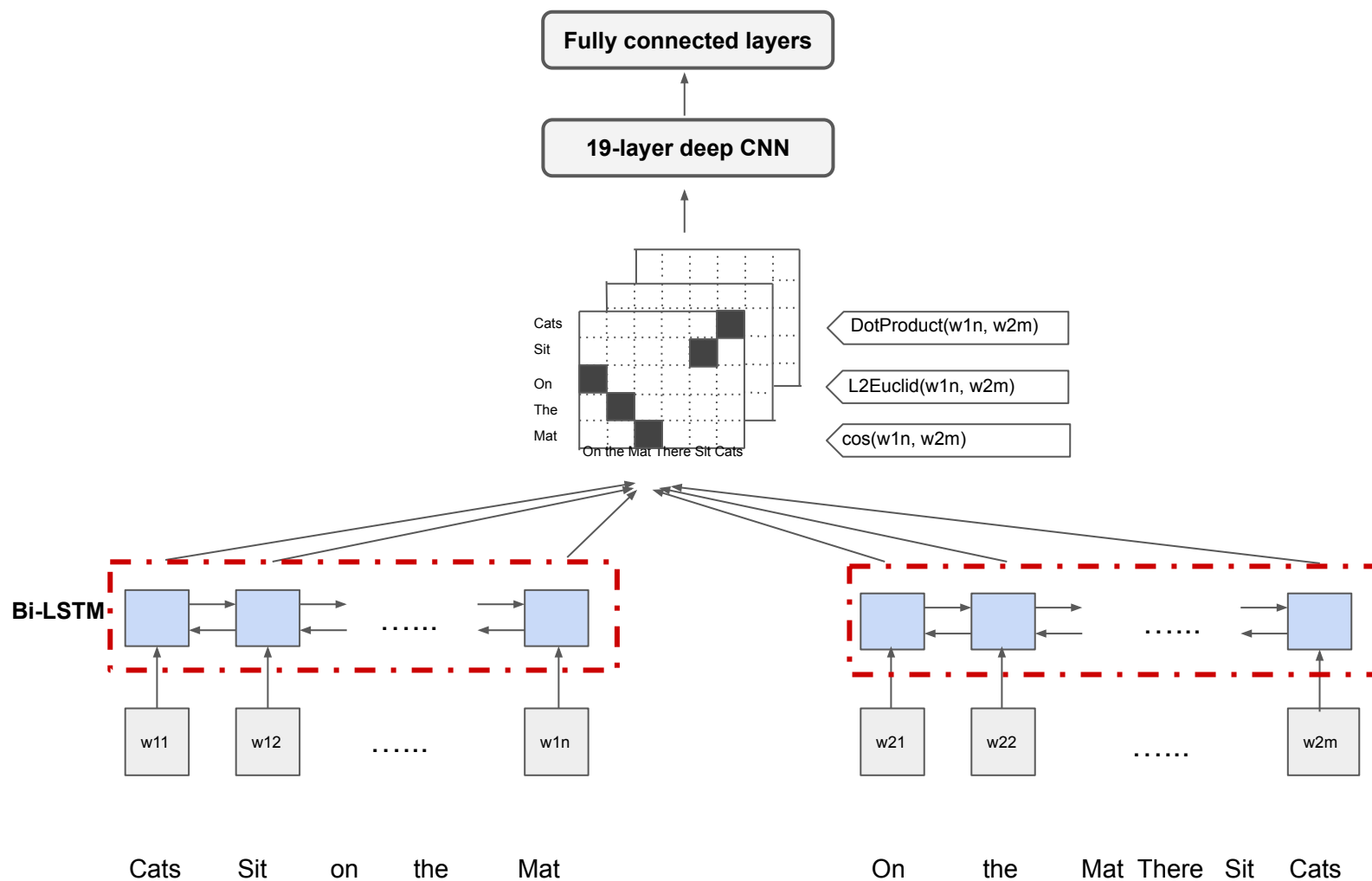


PWIM: He, Hua, and Jimmy Lin. "Pairwise word interaction modeling with deep neural networks for semantic similarity measurement." *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2016.

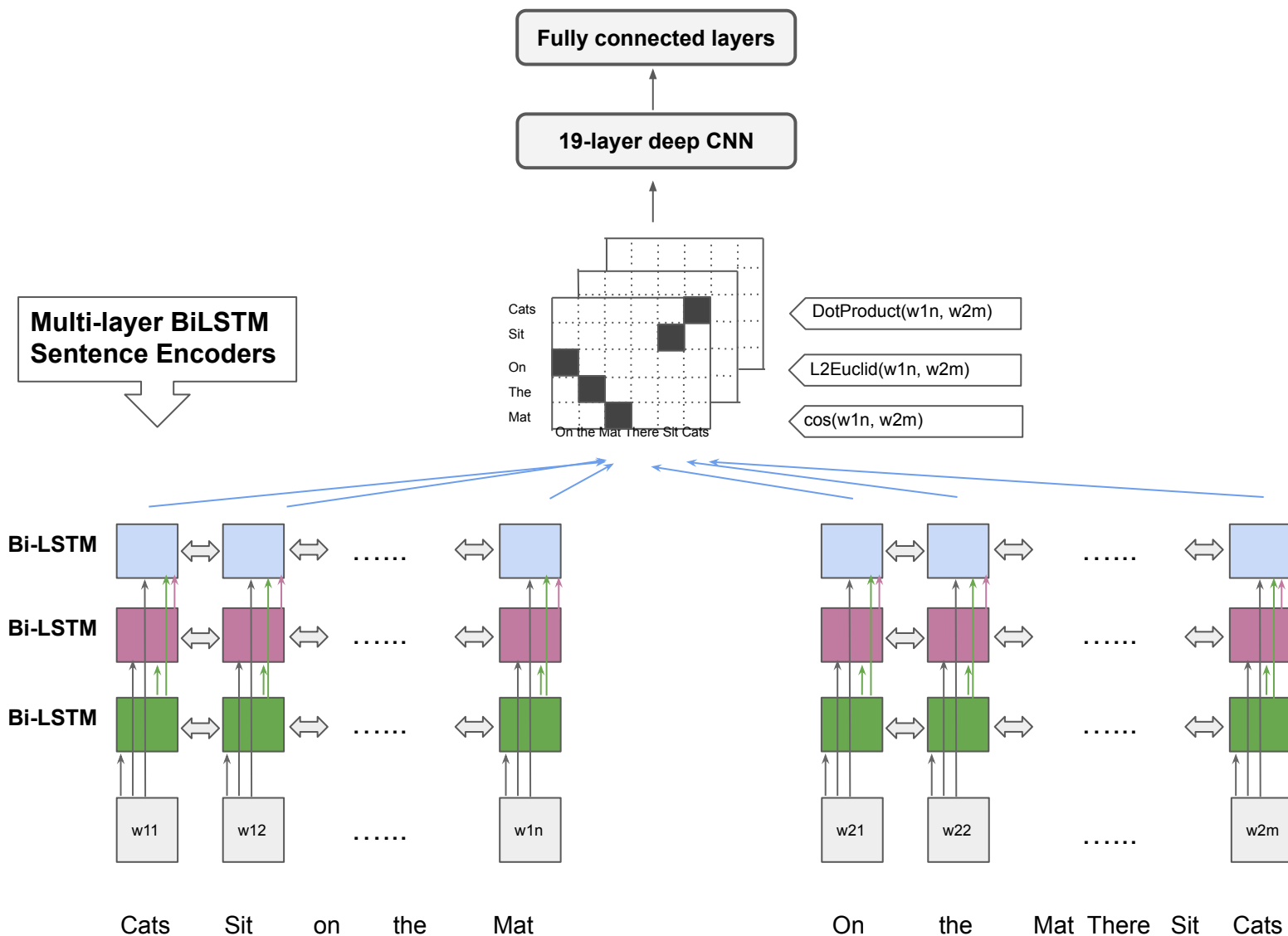
Pairwise Word Interaction Model (PWIM)



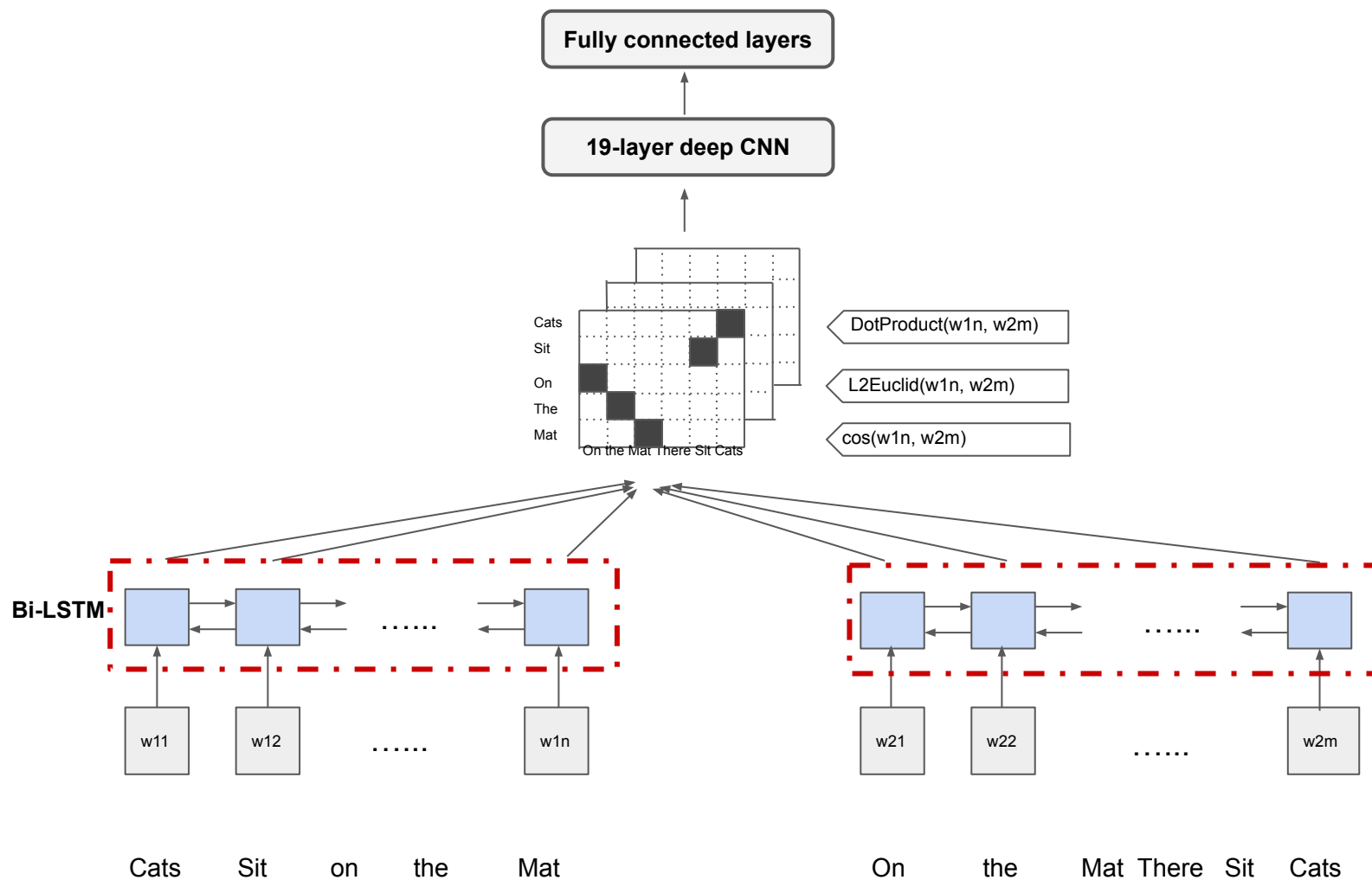
Pairwise Word Interaction Model (PWIM)



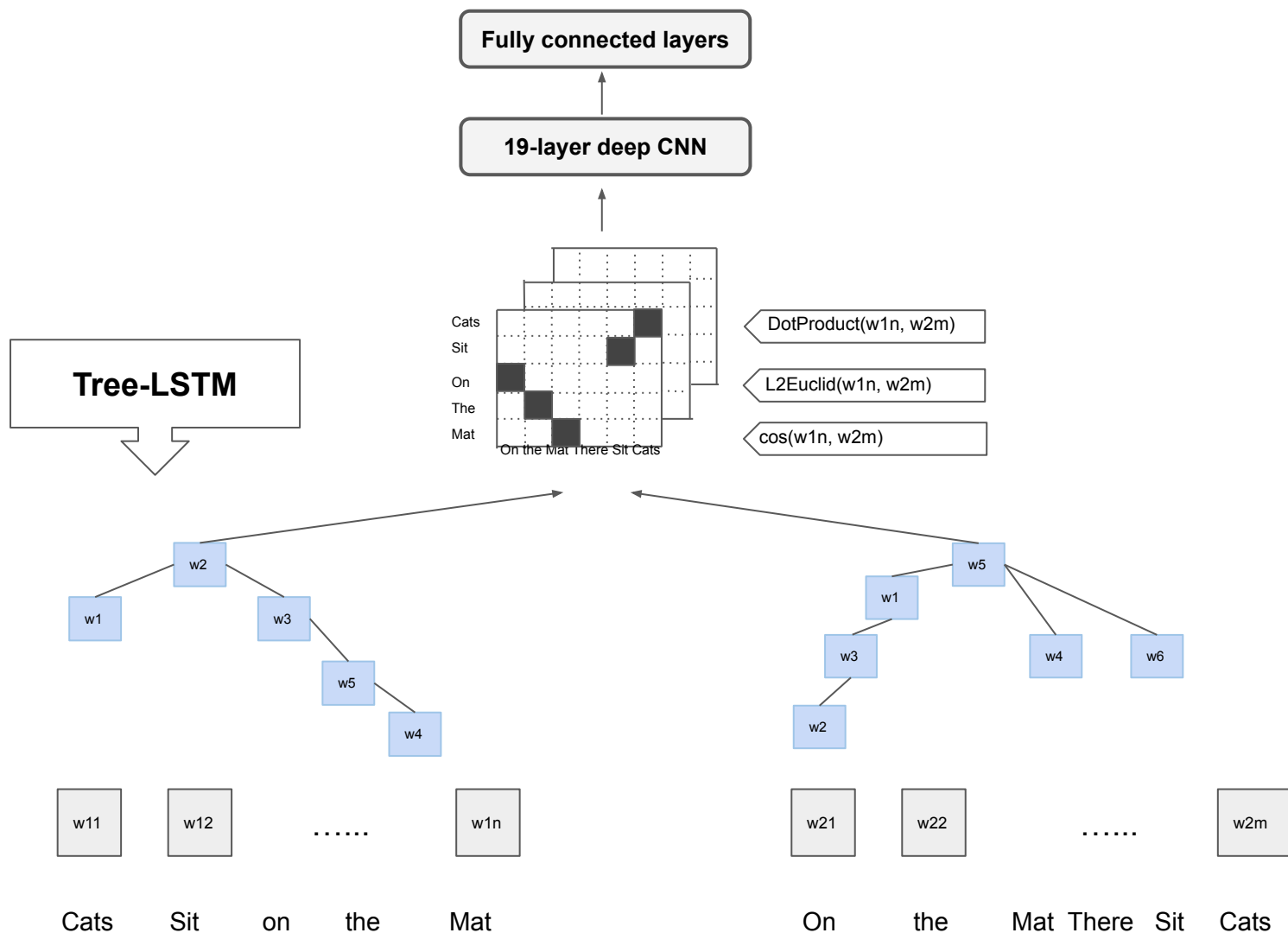
Pairwise Word Interaction Model (PWIM)



Pairwise Word Interaction Model (PWIM)



Pairwise Word Interaction Model (PWIM)



*Tai, Kai Sheng, Richard Socher, and Christopher D. Manning. "Improved semantic representations from tree-structured long short-term memory networks." *arXiv preprint arXiv:1503.00075*(2015).

Experiments on eight datasets

❑ Semantic Textual Similarity



[STS-2014](#) [SICK](#)

❑ Paraphrase Identification

● Paraphrase

● Non-paraphrase

[Quora](#) [Twitter](#) [PIT-2015](#)

❑ Question Answering

● True

● False

[WikiQA](#) [TrecQA](#)

❑ Natural Language Inference

● Entailment

● Neutral

● Contradict

[SNLI](#)

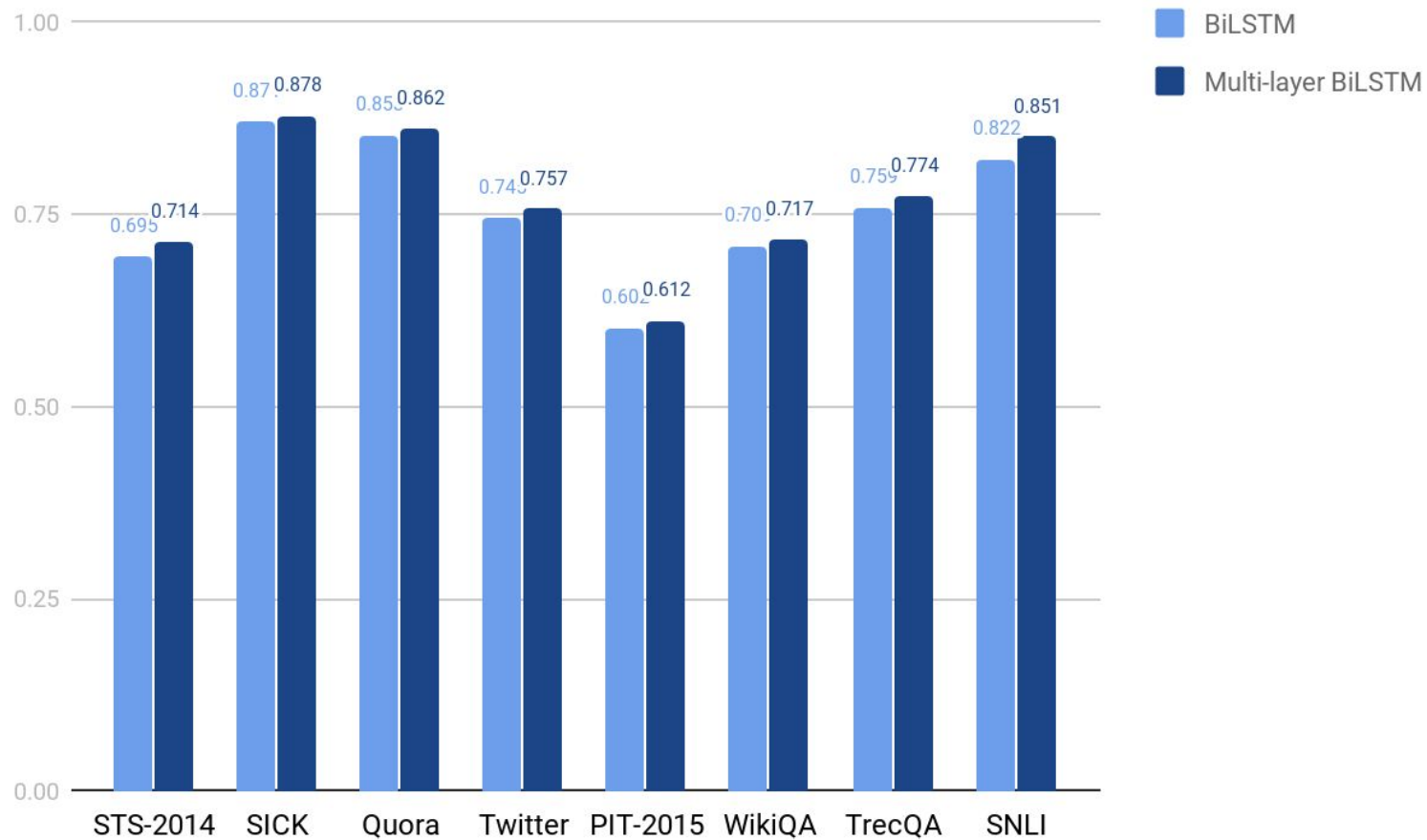
Experiment Results Analyses

Dataset	SNLI	Quora	Twitter	PIT-2015	STS-2014	WikiQA	TrecQA	SICK
	Acc	Acc	F1	F1	Pearson's r	MAP/MRR	MAP/MRR	Pearson's r/ ρ
InferSent	0.846	0.866	0.746	0.451	0.715	0.287/0.287	0.521/0.559	-
SSE	0.855	0.878	0.650	0.422	0.378	0.624/0.638	0.628/0.670	-
DecAtt	0.856	0.845	0.652	0.430	0.317	0.603/0.619	0.660/0.712	-
ESIM _{tree}	0.864	0.755	0.740	0.447	0.493	0.618/0.633	0.698/0.734	-
ESIM _{seq}	0.870	0.850	0.748	0.520	0.602	0.652/0.664	0.771/0.795	-
ESIM _{seq+tree}	0.871	0.854	0.759	0.538	0.589	0.647/0.658	0.749/0.768	-
PWIM _{our}	0.822	0.853	0.745	0.602	0.695	0.709/0.723	0.759/0.822	0.871/0.809
mPWIM _{seq}	0.851	0.862	0.757	0.612	0.714	0.717/0.728	0.774/ 0.835	0.878/0.821
mPWIM _{seq+tree}	0.855	0.870	0.743	0.623	0.718	0.735/0.751	0.781/0.821	0.887/0.834
Abs increase (%)	3.3	1.7	-	2.1	2.3	2.6/2.8	2.2/-	1.6/2.5

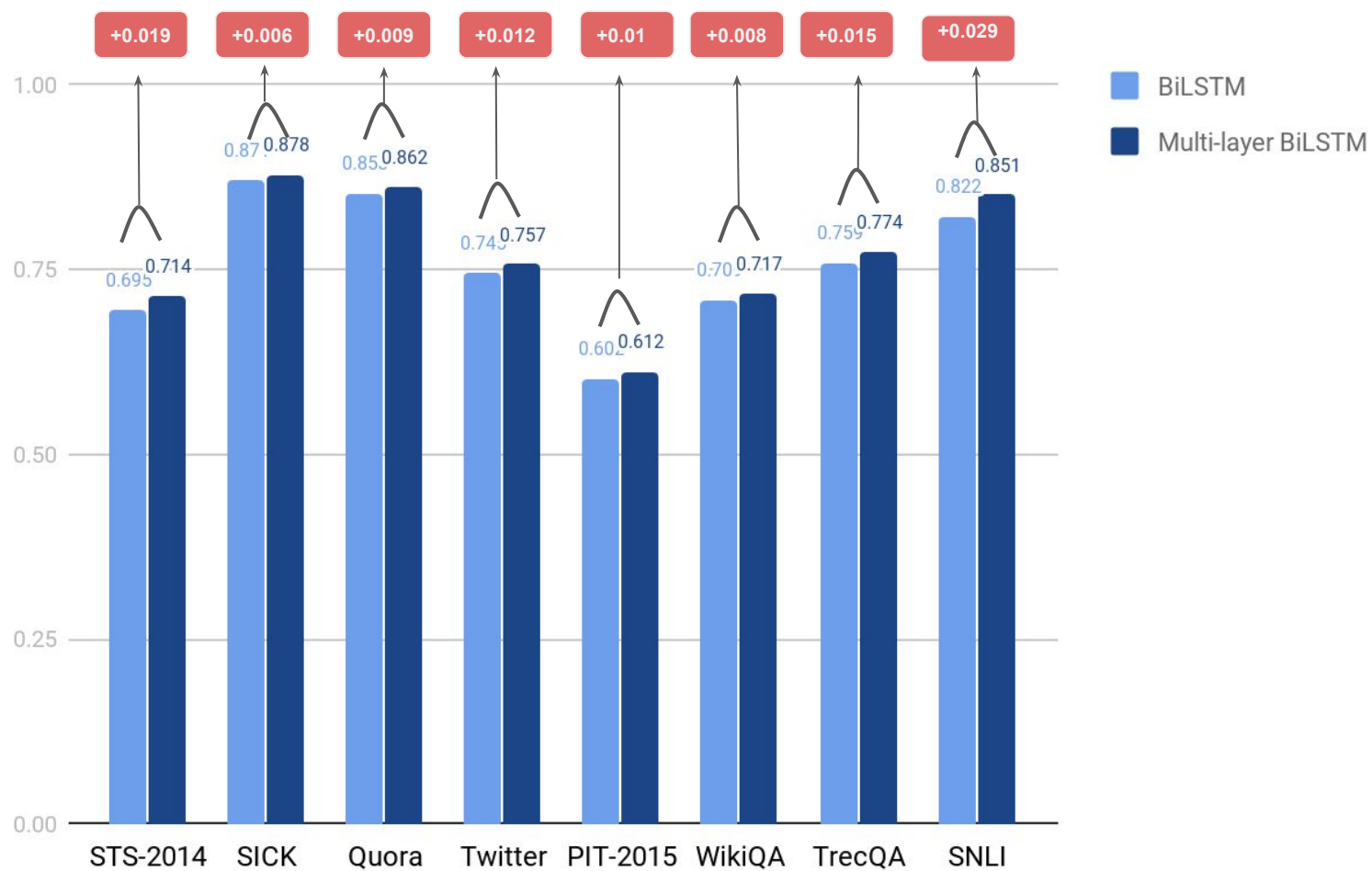


Ay,
caramba!

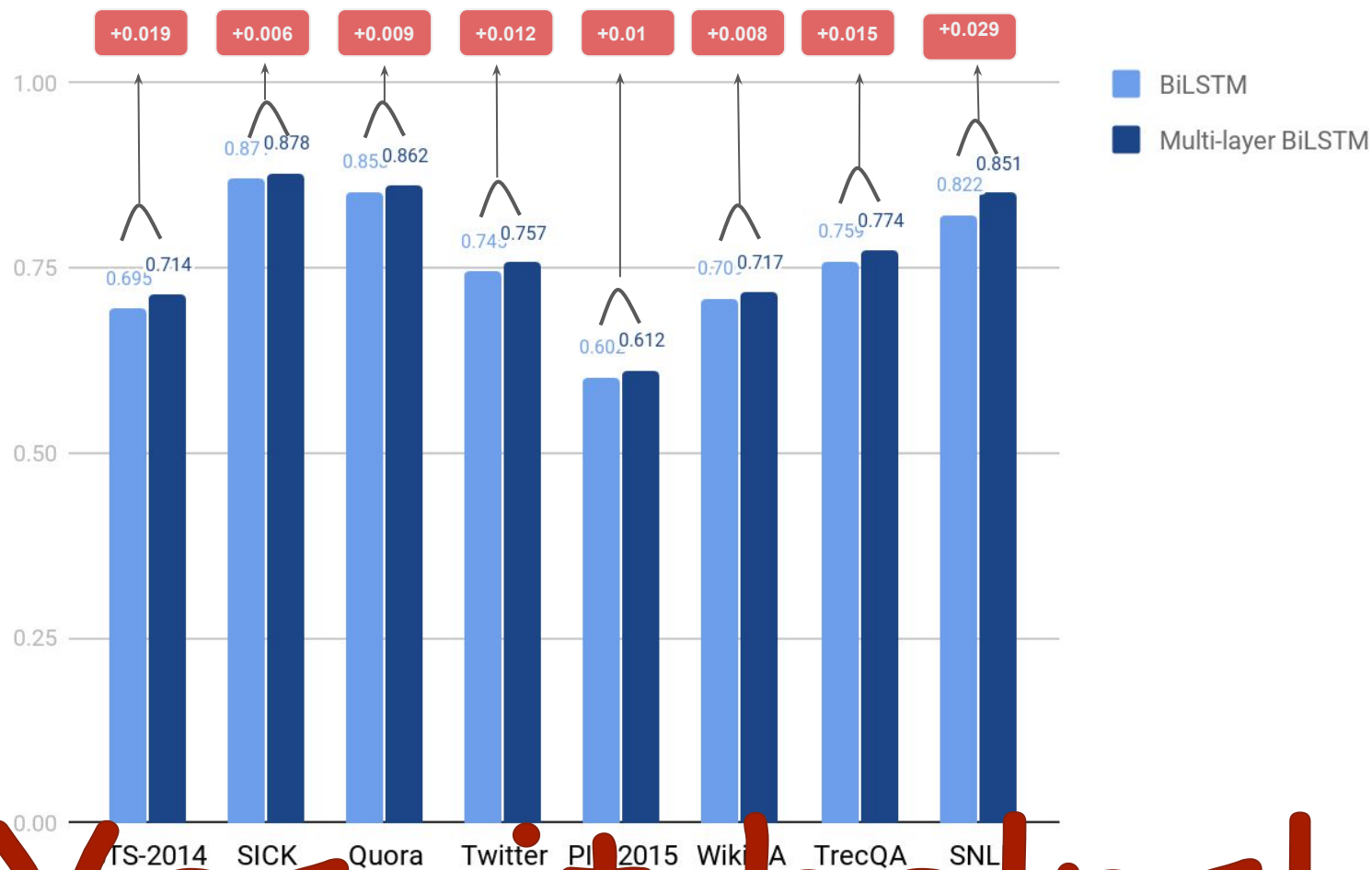
Additional Contextual Structure?



Additional Contextual Structure?

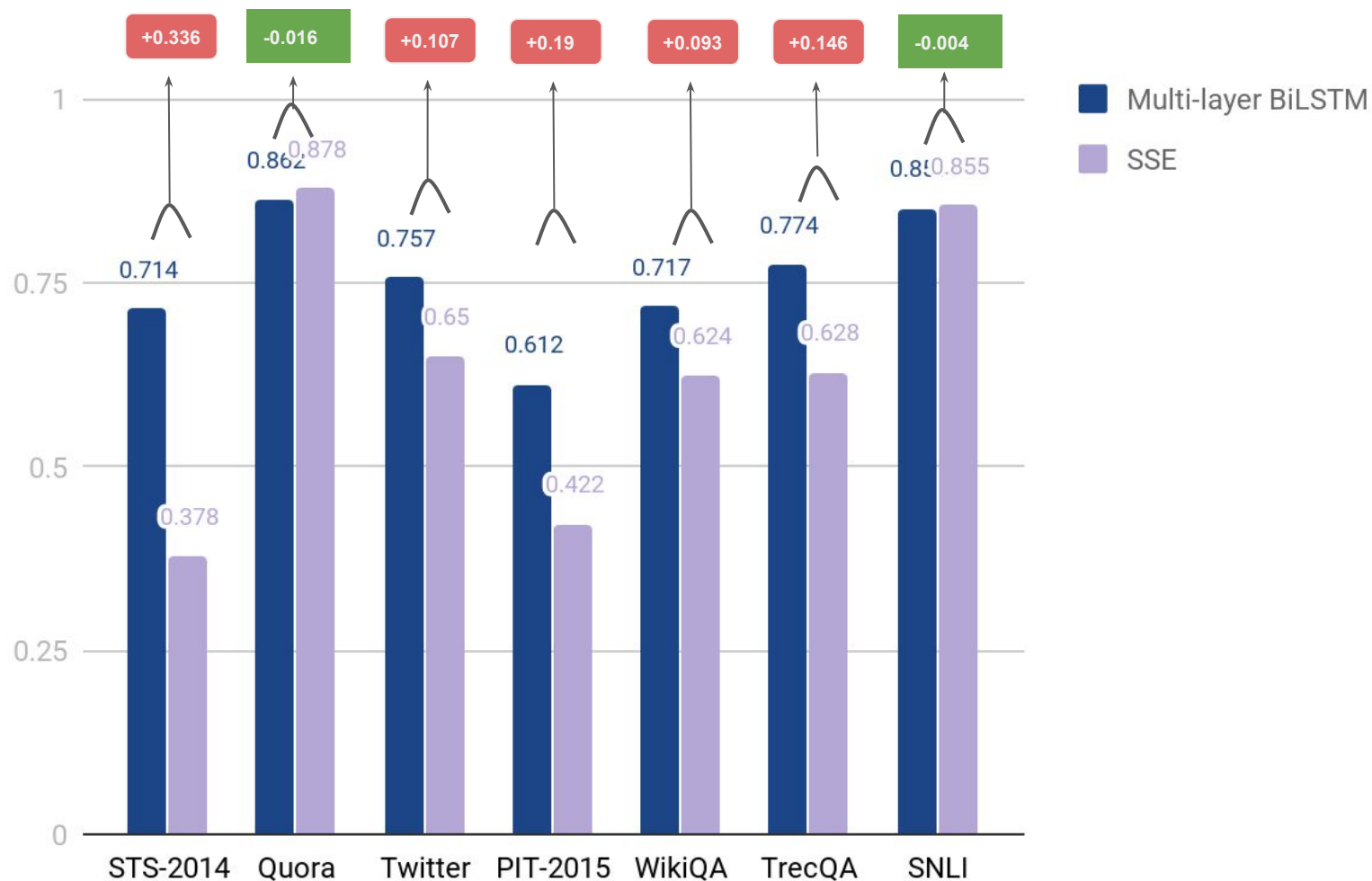


Additional Contextual Structure?

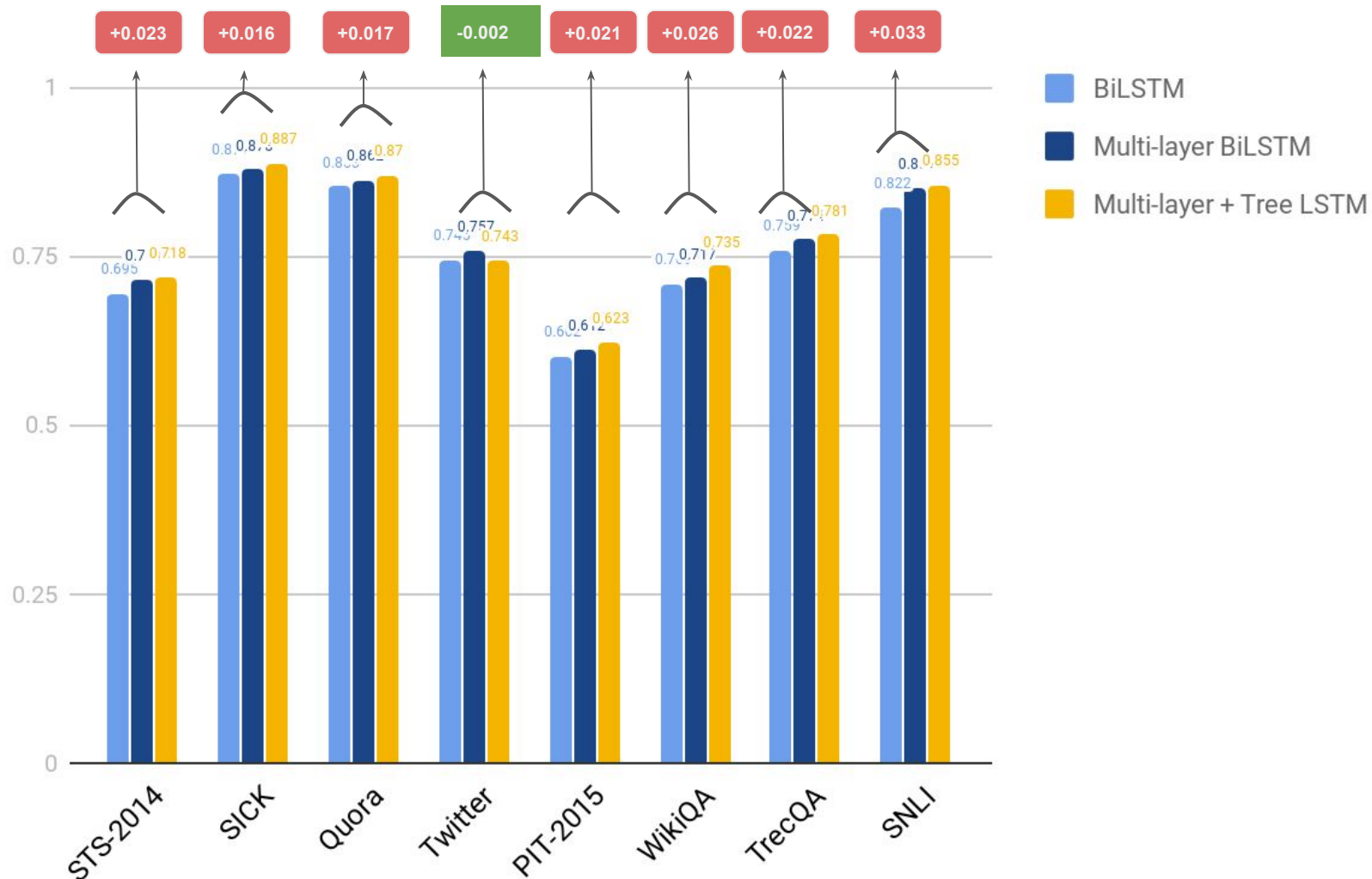


Yes, it helps!

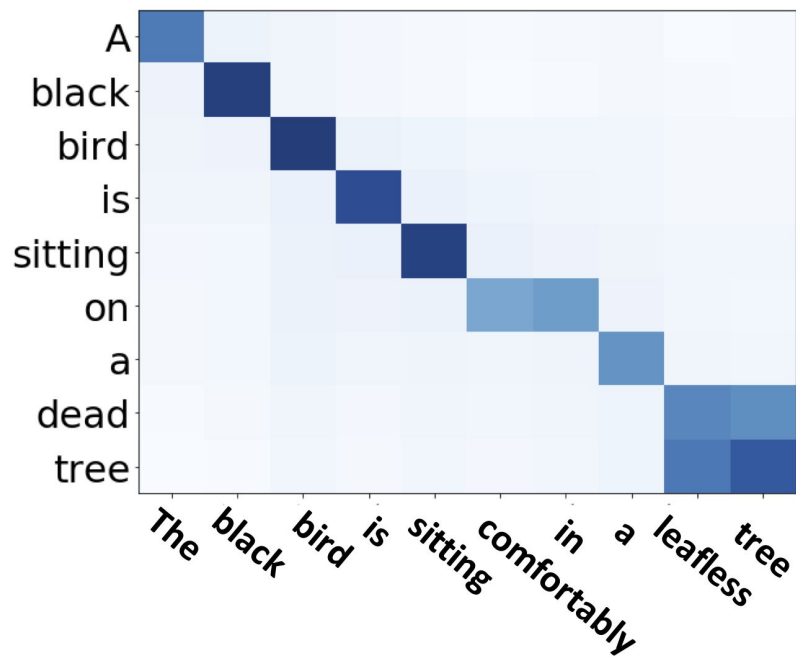
Additional Contextual Structure?



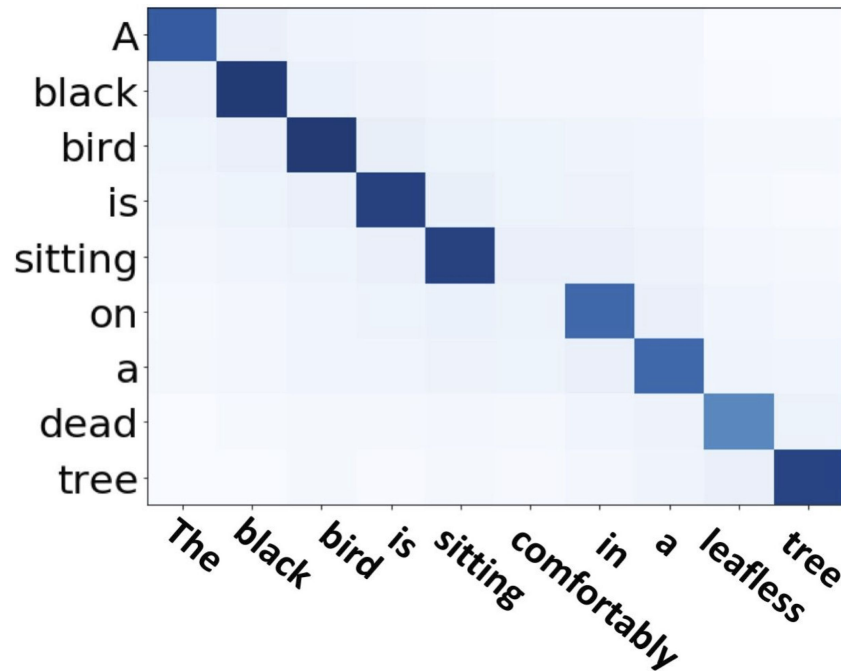
Additional Syntactic Structure?



Sentence Pair Visualization

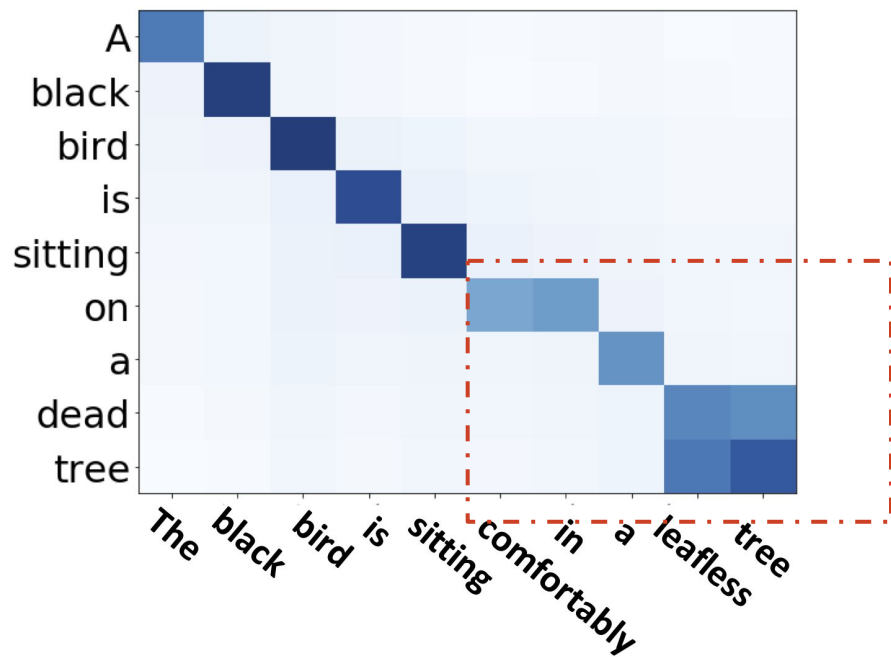


mPWIM_seq

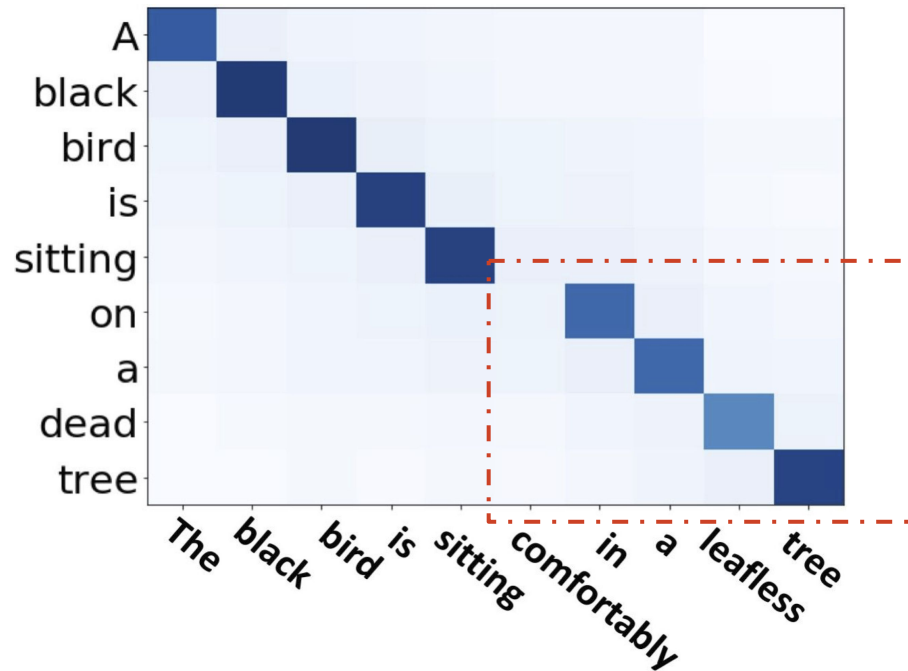


mPWIM_seq+tree

Sentence Pair Visualization

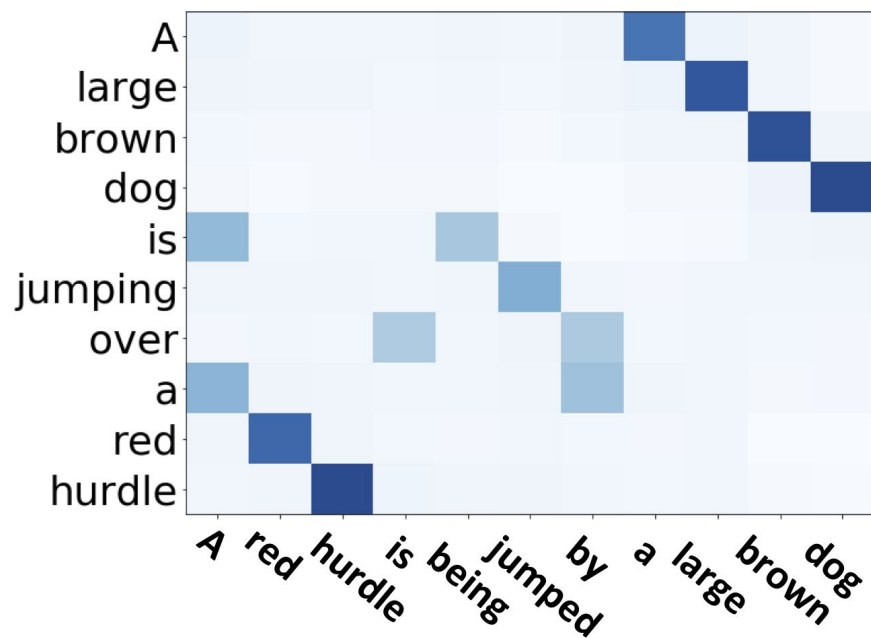


mPWIM_seq

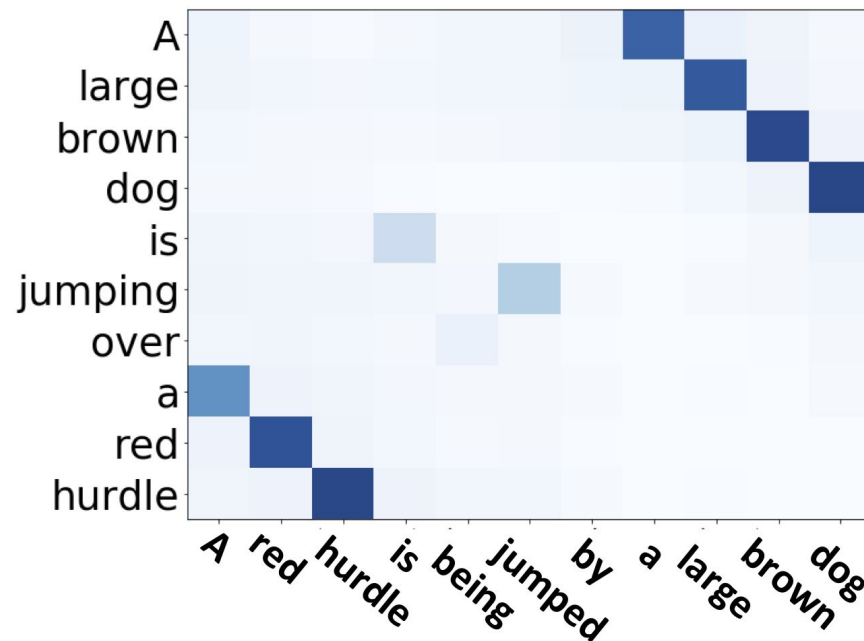


mPWIM_seq+tree

Sentence Pair Visualization

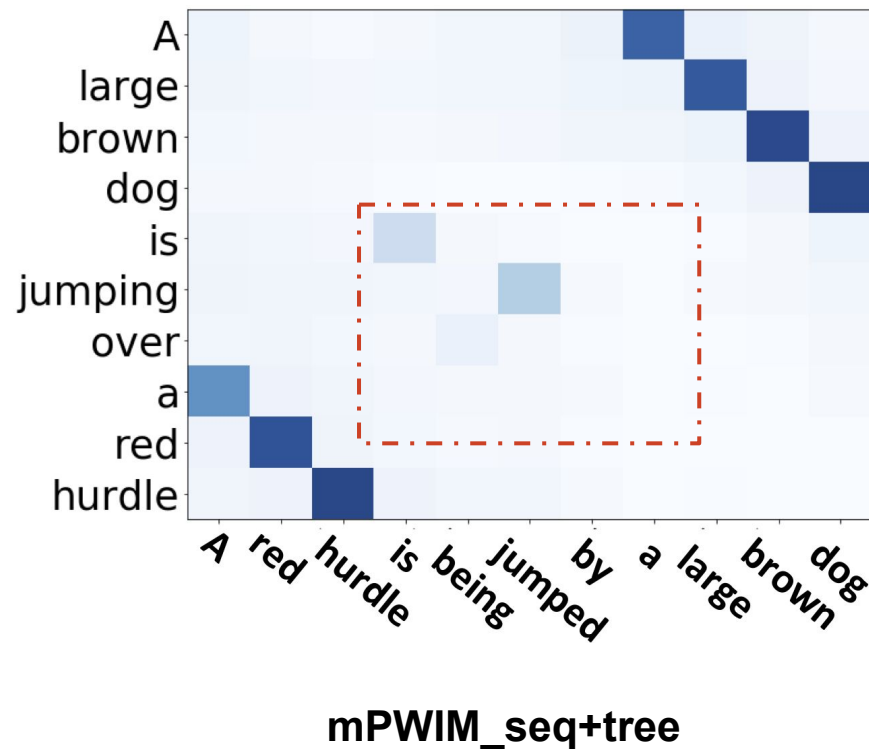
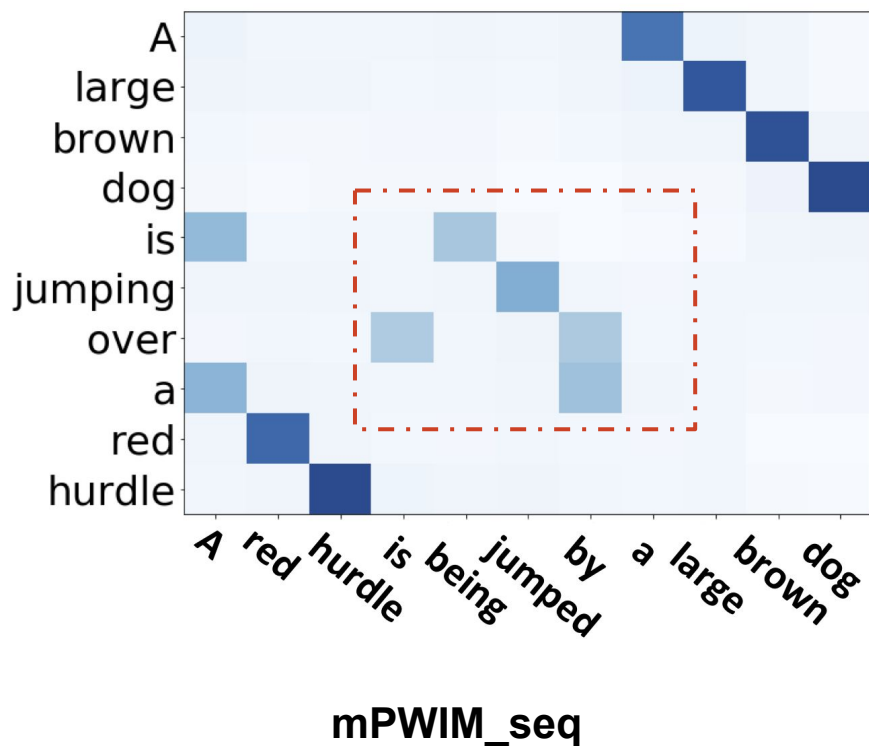


mPWIM_seq



mPWIM_seq+tree

Sentence Pair Visualization



What about BERT?



What about BERT?

- We are after the question: Does Contextual / Syntactic Structure help?
- Help to form a better student model



Takeaway

**Incorporating structural information
contributes to consistent improvements
over strong baselines**



Thanks for your attention!

